

## Weekly Report

### Done

1. 组会报告准备: Manifold: A Model-agnostic Framework for Interpretation and Diagnosis of Machine Learning Models

在阅读这篇文章时, 其中对于数据分布展示的设计与我之前做异常检测可视化时候很类似; 但是问题在于我们没有想到 model agnostic 这个概念, 也缺乏专家的真正反馈.

近年来, 在机器学习领域也有很多 model agnostic 的 interpretation 方法, 相当于是给模型开箱, 但并不与模型本身架构相关(似乎由 16 年一篇 KDD 文章 LIME 带动); 今年 VIS 有一篇 RuleMatrix, 根据复杂模型到处简单的 rules, 并可视化. 这篇文章设计目标是那些领域专家但是非 ML 专家, 这个针对点选的比较合适. 光从 rule 提取可视化这个点其实之前也有想过, 但最终没有下手.

Manifold 又是把模型的 box 给关上了, 从不同模型的共同点出发, 进行模型诊断和调理.

2. 关于各种 XX 学习支持可视化方面, 去年 vis 有一篇简单的关于数据可视化 visual saliency 模型的论文, 是结合了传统图像中 saliency 以及文字 saliency 的模型. 我有一个想法是, 能不能模拟马哥散点图一文的做法, 做一些针对特定可视化形式的 visual saliency 模型?

或者更发散一点, 这次 vis 有不少大屏小屏手机屏可视化方面的; 就以一个 infographics 的电脑端和手机端而言, 他们也往往都是要重新制作, 否则电脑上好看的手机上完全不能看. 这可能也是一个不同尺寸下可视化的 visual saliency 不一致的问题.

3. 关于 visual report 生成, 和元哲聊聊他一些想法(根据数据随机生成一种 view 的布局), 但是评估起来有困难(如一个布局的好坏), 以及本身设计实力上也不够, 想法比较理想而不易实现. 但如果像我那样又限定在某个场景中, 那可扩展性又不强, 有点矛盾.
4. 关于几种 fusion 的问题, 在于陈老师讨论后有了新的理解, 也应该是更贴近马老师原本想法的. 我找到了 infovis 的 reference model, 其中有 data transformation, visual mapping 与 view transformation 三个环节; 我们在前两个环节做文章.

在有关 graph 的文章里, NodeTrix 是比较容易读懂的, 而 Interactive Level-of-Detail Rendering of Large Graphs 和其前作因为涉及到了一些图形学知识只粗略了解了其设计含义. NodeTrix 生成仍是一个需要交互与知识的, 且其 case 用的数据量相对而言也不大.

NodeTrix 既可以从纯 node-link 出发交互构建也可以从纯 matrix 出发构建. 在数据量很大时, 两个方式都会吃力. NodeTrix 一开始是为了便于分析 graph 中的 communities; 这个可能可以和上面的 LoD 结合起来, 但我们可能是从数据出发去寻找 communities?

在我了解范围内, 除了 LoD, graph 的交互探索还有像 DoI 技术. LoD 本身, 梅博还提到了 semantic zooming 的交互. 正如我们那天讨论的, 是否可以在有多重属性的图上做类似的不同 fusion 的展示. 这些下周继续思考和讨论.

### To Do

1. 详细问问巫老师那边对于 transfer learning 的见解. 尽管上次也提到 Draco, 但点有点太大, 看看能不能缩小范围.
2. 上面 2, 4 中提到的讨论. 2 也许已经有人在做, 4 我需要先再熟悉下论文再去问问郭博或者东明有没有实际操作.

## 论文阅读

*Draco: Formalizing Visualization Design Knowledge as Constraints* 仔细阅读理解了其中的逻辑语言、软 constraint 和硬 constraint 的方式, 以及构建回归模型的数据准备方式. 对于想要自动生成可视分析系统而言, 可能要求过高了; 但是如果是简单图表构成的 dashboard, 我觉得是可期的, 尽管当前 Draco 只针对单个图表. 另外, 注意到其训练模型时, 用的是之前 *EuroVis Assessing effects of task and data distribution on the effectiveness of visual encodings* 里面收集的众包 user study 结果.

其余文章见上面讨论.

## 工作时间

平时 10, 周末一共 6 小时. 总共 56.

## 个人规划

性质	Ddl	进展	目标
中长期	2019 年中	统计: 假设检验导论	统计 优化 算法
短期	11.17	暂无 (转为对 fusion 讨论)	Visual report
短期	18 年底	暂无	XX 学习对可视化支持
长期			博士毕业论文中的研究方向